Do Conjunction Errors in Auditory Recognition Imply Feature Migration?

Randall K. Jamieson, Queen's University

William Forde Thompson, University of Toronto

Lola L. Cuddy and D. J. K. Mewhort, Queen's University

Abstract It has recently been proposed that the feature values (e.g., pitch, duration, etc.) of tones are processed in separate perceptual channels and are later reintegrated to form the experience of a unitary percept. Patterns of errors in recognizing combinations of auditory features have been taken as failures of, and thus evidence for, an auditory feature-integration process. We outline an alternative interpretation of these effects based on similarity structure of the stimuli.

When observers are presented with several visual objects, but are required to attend to another position in space, they may report the features from the objects in mismatched combinations. For example, when presented with a red square and a blue circle, observers may report having seen a red circle, an *illusory conjunction* obtained by mismatching colour with shape (Treisman, 1986; Treisman & Gelade, 1980; Treisman & Paterson, 1984; Treisman & Schmidt, 1982).

Illusory conjunctions can be explained in terms of feature-integration theory. According to feature-integration theory, primitive attributes of stimuli, called *features*, are extracted preattentively and are later reintegrated, or bound, to form a unitary percept. The binding process is thought to be error prone, particularly when attention is taxed.

Feature-integration theory has been developed in studies of visual perception (e.g., Treisman, 1986; Treisman & Gelade, 1980; Treisman & Schmidt, 1982). However, the theory has also been applied to auditory perception (Deutsch, 1986; Hall, Pastore, Acker, & Huang, 2000; Thompson, 1994; Thompson, Hall, & Pressing, 2001; Thompson & Sinclair, 1993). Thompson (1994), for example, presented two repeating sequential tones that differed in pitch and duration. Listeners were required to detect changes to the pattern. Changes included (1) a new pitch or duration, a *single* change, (2) a new pitch and a new duration, a *double* change, and (3) a re-combination of the pitch from one tone with the duration of the other, a *switch* change. Single and double changes were easier to detect than switch changes, and detection of switch changes worsened when listeners were distracted.

Feature-integration theory instructs that detection of single and double changes requires only accurate feature registration whereas detection of switch changes requires both accurate feature registration and accurate feature binding. Thus, feature-integration theory explains the extra difficulty in identifying switched features as failures of feature binding. The explanation carries feature-integration theory over from vision to audition.

A second class of evidence for auditory feature-integration theory comes from paired-feature auditory recognition-memory studies (Thompson et al., 2001). In Thompson et al.'s experiments, listeners studied sets of sequentially presented tones that varied on two stimulus attributes: pitch and duration. Each list was followed by a probe stimulus and listeners were required to indicate whether or not the probe matched a tone in the studied set. A match response to a positive probe (an item from the list) was a hit, and a match response to a negative probe (an item not in the list) was a false alarm.

Negative probes differed from the study tones in one of four ways: 1) One kind of probe combined the pitch from one tone with the duration from another, a *switch* change; 2) A second kind combined the pitch from one tone with an unstudied duration, a *single-1* change; 3) The third kind combined the duration from one tone with an unstudied pitch, a *single-2* change; and 4) The fourth combined an unstudied pitch with an unstudied duration, a *double* change.

Thompson et al. (2001) reported more match responses for switch probes than for the other three kinds of negative probes. Of particular significance,

there were more match responses for switch probes than for single probes and the match response rate for switch probes was close to that for positive probes. Thompson et al. used a multinomial model and the error rates for the five kinds of probes to estimate the probabilities of misperceiving simple features and to estimate the probability of illusory conjunction errors. Based on their analysis they suggested that the pattern of match response rates for the five conditions indicates that features were weakly or improperly integrated at the perceptual stage at which pitch and duration are initially conjoined. In a second experiment, Thompson et al. repeated pitches and durations across seven tones of a study set. The results were consistent with the first experiment and with the interpretation based on feature-integration theory.

Insensitivity to combinations of musical parameters is relevant to issues in music cognition and auditory perception. It likely influences music experience and, for that reason, has implications for music composition, for the perception of musical structure, and for the perception of musical similarity. The results are also important as evidence for auditory illusory conjunctions because they extend feature-integration theory's application from the visual-spatial domain to the auditory-temporal domain. Finally, because illusory conjunctions occured when tones were presented sequentially, and because feature-integration theory specifies that an illusory conjunction requires the concurrent availability of features from studied stimuli so they can be incorrectly bound, Thompson et al.'s (2001) results point to the importance of a mechanism that maintains features of a tone as free-floating perceptual units both over time and prior to binding.

In this report, we suggest an alternative interpretation of Thompson et al.'s (2001) experiments based on the role of similarity structure in their experiments. To outline the approach, consider a simple count of the features shared by a study set and a probe in Thompson et al.'s experiments. Positive probes shared both pitch and duration with a studied item. The pitch and duration of a switch negative probe shared corresponding features in the study set, but their conjunction did not. The pitch of a single-1 probe shared a pitch from the study set, but its duration did not. The duration of a single-2 probe shared a duration in the study set but its pitch did not. There were no features shared between a double probe and items in the study set. Because similarity facilitates the recognition of a positive probe and hinders the rejection of a negative probe, the similarity structures in the experiments anticipate the pattern of results predicted from feature-integration theory: $p(Match \mid \text{Positive}) > p(Match \mid \text{Switch}) > p(Match \mid \text{Single}) > p(Match \mid \text{Double})$.

Global recognition-memory theory (Clark & Gronlund, 1996; Humphreys, Pike, Bain, & Tehan, 1989) bases predictions about recognition performance on the degree of similarity between a probe and studied items. We demonstrate that similarity structure predicts the rank order of results observed by Thompson et al. (2001), and apply a contemporary global recognition-memory model, REM (Shiffrin & Steyvers, 1997), to formalize the argument.

Simulations Based on Similarity Structure

Global recognition-memory theory postulates that studied items are stored in memory and that each encounter with an item is stored as a separate trace. During recognition, the probe accesses the traces of all studied items in parallel. Each trace activates according to its similarity with the probe, and the summed activation across traces indexes the likelihood that the probe will be recognized (Clark & Gronlund, 1996; Humphreys, Pike, Bain, & Tehan, 1989). The more similar the probe is to studied items, the more likely that a match response is given.

Global recognition-memory theory expresses the probability of a match response as a function of the similarity between a probe and studied items (Humphreys et al., 1989). Similarity takes a value, $\tau2$, if a probe and a study item have two features in common, a second value, $\tau1$, if they have one feature in common, and a third value, $\tau0$, if they have no features in common. The probabilities of a match response for each kind of probe that Thompson et al. (2001) used in their tasks can be written as a linear combination of $\tau2$, $\tau1$, and $\tau0$ – where $\tau2 > \tau1 > \tau0$ and N is the number of items in the study set:

$p(Match \mid \text{Positive}) = \tau2 + (N - 1)\, \tau0$
$p(Match \mid \text{Switch}) = 2\, \tau1 + (N - 2)\, \tau0$
$p(Match \mid \text{Single}) = \tau1 + (N - 1)\, \tau0$
$p(Match \mid \text{Double}) = N\, \tau0$

As shown in Humphreys et al.'s (1989) equations above, similarity structure anticipates that switch probes are more likely to elicit a match response than single probes on the basis of similarity, and independent of the opportunity for feature migration. Recognition memory has given rise to several formal models, including REM (Shiffrin & Steyvers, 1997), Minerva2 (Hintzman, 1984, 1986, 1988), TODAM (Murdock, 1982, 1983), CHARM (Eich, 1982), and SAM (Raaijmakers & Shiffrin, 1981). The models are referred to collectively as global matching recognition-memory models.

REM (Shiffrin & Steyvers, 1997) is a contemporary global recognition-memory model that calculates simi-

larity between the probe and memory using a Bayesian log odds ratio – the likelihood of an observed match to memory given either a target or a foil. Mean familiarity across all traces in memory indexes the likelihood that a match response is elicited. We applied REM to the paired-feature recognition tasks used by Thompson et al. (2001).

Simulations With REM

Stimuli were represented at two levels: Features (i.e., pitch and duration values) and tones (i.e., combinations of pitch and duration). Feature vectors of 10 elements represented pitch and duration values. Each vector element was assigned a randomly sampled value from a geometric distribution, with the p(success) = g. Tone vectors were created by concatenating two feature vectors: The first 10 elements represented the tone's pitch, the second 10 elements represented its duration, and the pairing represented the conjunction of features.

Recognition probes were represented as tone vectors. Positive probes were copies of studied tones. For switch probes, the first 10 elements from one study tone were concatenated with the second 10 elements of another (i.e., the pitch value from one tone with the duration value from another). For single-1 probes, the first 10 elements from one study tone (i.e., the pitch feature) were concatenated with a new feature vector. For single-2 probes, a new feature vector was concatenated with the second 10 elements (i.e., the duration feature) of a study tone. For double probes, two new feature vectors were generated and concatenated.

For each simulated trial, the program generated a set of study items and a corresponding set of recognition probes (a match, a switch, a single-1, a single-2, and a double probe). The study-items were encoded into a memory matrix. Encoding involved three parameters: 1) the number of opportunities to encode each element, R; 2) the probability that an element is noticed, u^*; and 3) the probability that a noticed feature is encoded correctly, c. A noticed and correctly copied element took its value from the corresponding stimulus representation. A noticed, but incorrectly copied element, took a new, random value. Once an element was noticed and encoded, it was not overwritten. Elements that went unnoticed during encoding were assigned a value of zero. Feature conjunctions in the study tones were preserved in memory, blocking feature migration. For all our simulations, R was set to 1 and u^* was set to .5.

After the study tones were encoded, similarity for each kind of probe was calculated against tones in the memory matrix as a Bayesian odds ratio (Shiffrin & Steyvers, 1997, p. 147),
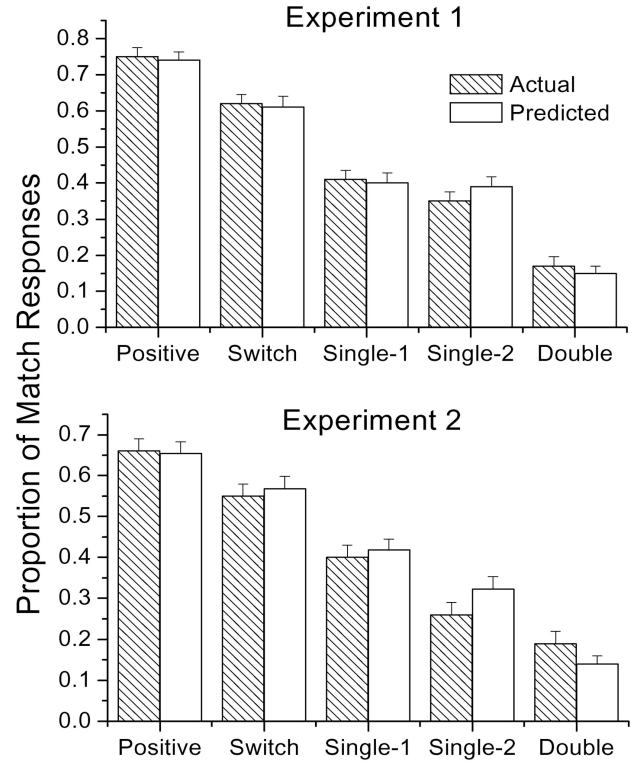

Figure 1. Mean proportions of match responses for each kind of recognition probe. Error bars represent standard errors. The top panel shows the actual and simulated data for Experiment 1, and the bottom panel shows the actual and simulated data for Experiment 2.

$$\lambda_t = (1-c)^{n_{tq}} \prod_{i=1}^{\infty} \left[\frac{c + (1-c)g(1-g)^{i-1}}{g(1-g)^{i-1}} \right]^{n_{itm}} ,$$

where λ_t is an index of similarity between the probe and a study tone t, n_{tq} is the number of mismatching nonzero values between the probe and the study tone, and n_{itm} is the number of matching nonzero values between the probe and the study tone that have value i.

Mean familiarity for the probe, F_p, was calculated by taking the $\log_{10}$ value of the average λ_t value across all N probe-study tone comparisons,

$$F_p = \log_{10}\left(\sum_{t=1}^{N} \lambda_t / N \right) .$$

A match response was counted when familiarity was equal to or greater than a decision criterion, k. The $\log_{10}$ of mean similarity is used because similarity values, λ_t, are highly skewed. By taking the $\log_{10}$ of similarity, values retain their differences but are spread on a more intuitively meaningful scale.

Simulations for Experiment 1

In Thompson et al.'s (2001) first experiment, listeners studied sets of two tones that differed in both pitch and duration and were required to judge whether or not a probe was in the studied set.

Each simulation comprised 320 recognition trials – equivalent to the number of trials contributed by the 20 participants in Thompson et al.'s experiment. For each trial, the program generated a set of two stimulus-items and five recognition probes, encoded the stimulus-items, and calculated the familiarity for each probe against the encoded items. When familiarity exceeded a decision criterion, k, a match response was recorded. Proportions of match responses were scored by dividing the number of match responses for each kind of probe by the number of simulated trials. Because simulations had the same number of trials as the experiment, standard errors reported for the five means are comparable to those reported by Thompson et al.

Predicted values are means taken from 10 simulations, each comprising 320 recognition trials. A downhill Simplex algorithm was used to search the three-dimensional free-parameter space that includes c, k, and g (see Press, Teukolsky, Vetterling, & Flannery, 1992, pp. 402-406). Predicted and actual proportions of match responses are reported in the upper panel of Figure 1 as a function of correct encoding, $c = 0.72$, the decision criterion, $k = 0.14$, and the shape of the geometric distribution that items were sampled from, $g = 0.87$.

Predictions matched the rank order of match response rates predicted from feature-integration theory: $p(Match \mid Positive) > p(Match \mid Switch) > p(Match \mid Single) > p(Match \mid Double)$, $RMS = .021$. The higher match response rate for switch probes compared to single probes, which has been taken to imply feature migration, falls out from the similarity structure of the experiment.

Other global recognition-memory models, like Minerva2 (Hintzman, 1984, 1986, 1988) and TODAM (Murdock, 1982, 1983), yield the same outcomes as REM (Shiffrin & Steyvers, 1997). The shared ability to predict the pattern of match responses emerges because global recognition-memory models base recognition predictions on the similarity between a probe and a set of studied items.

Simulations for Experiment 2

In Thompson et al.'s (2001) second experiment, study sets comprised seven tones made from combinations of two durations and six pitches. The first and last tones in a sequence were identical. Tones in positions 2 through 5 of a sequence each had a different pitch. Two of those tones shared one duration and the remaining three tones shared another duration (see p. 132, Thompson et al.).

Studied tones were represented as vectors comprising two 10-element fields, one representing pitch and the other duration, with concatenation representing feature pairings. Negative probes were constructed by adding unstudied fields and rearranging studied fields as required. In our simulations, positive probes were copies of a study tone comprising the less frequent duration. Switch probes comprised a studied pitch mismatched with the less frequent duration. Single-1 probes comprised the pitch from one study tone paired with a new duration. Single-2 probes comprised a new pitch paired with the less frequent duration. Double probes comprised an unstudied pitch value and an unstudied duration value.

Each simulation comprised 288 recognition trials – equivalent to the number of trials contributed by the 18 participants in Thompson et al.'s (2001) experiment. On each trial, the program generated a study set and a corresponding probe set. The study set was encoded and each kind of probe was compared against it. Familiarity was calculated for each probe. When mean familiarity across the seven traces was equal to or greater than a decision criterion, k, a match response was counted. Proportions of match responses were scored by dividing the number of match responses for each kind of probe by the number of simulated trials. Because simulations had the same number of trials as the experiment, standard errors reported for the five means are comparable to those reported by Thompson et al.

Predicted values are means taken from 10 simulations, each comprising 288 recognition trials. Predicted and actual match response rates are reported in the lower panel of Figure 1. We re-used the fitted parameters from the first simulation: $c = 0.72$, $k = 0.17$, and $g = 0.87$.

Using the same parameter values from the first simulation, the results matched the rank order of match response rates predicted by feature-integration theory: $p(Match \mid Positive) > p(Match \mid Switch) > p(Match \mid Single) > p(Match \mid Double)$, $RMS = .037$. The higher match response rate for switch probes compared to single probes, which appears to imply feature migration, falls out from the similarity structure of the experiment.

Conclusion

Thompson et al. (2001) used a recognition-memory task to argue for feature migration in the auditory domain. We have outlined an alternative explanation

based on similarity structure formalized using current theory from the recognition-memory literature. Our work shows that similarity structure produces the pattern of match responses previously attributed to feature migration.

Our analysis of the role of similarity exploits a model for recognition-memory, and we have restricted our conclusions to tasks that contain a prima-facie memory component. Other examinations of auditory feature migration have minimized the role of memory (e.g., Hall et al., 2000; Thompson, 1994). However, even in tasks that minimize the role of memory, similarity relations between a probe and a stimulus-array remain a potential influence on performance.

Randall K. Jamieson, Department of Psychology, Queen's University, Kingston, Ontario, Canada; William F. Thompson, Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario, Canada; Lola L. Cuddy, Department of Psychology, Queen's University, Kingston, Ontario, Canada; D. J. K. Mewhort, Department of Psychology, Queen's University, Kingston, Ontario, Canada.

Correspondence concerning this article should be addressed to Randall K. Jamieson, Department of Psychology, Queen's University, Kingston, Ontario K7L 3N6. (E-mail: randy@nets.psyc.queensu.ca).

References

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37-60.

Deutsch, D. (1986). Auditory pattern recognition. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 1-49). New York: Wiley.

Eich, J. M. (1982). A composite holographic associative recall model. *Psychological Review, 89,* 627-661.

Hall, M. D., Pastore, R. E., Acker, B. E., & Huang, W. (2000). Evidence for auditory feature-integration with spatially distributed items. *Perception & Psychophysics, 62,* 1243-1257.

Hintzman, D. L. (1984). MINERVA2: A simulation model of human memory. *Behavior Research Methods, Instruments & Computers, 16,* 96-101.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411-428.

Hintzman, D. L. (1988). Judgments of frequency and recognition-memory in a multiple-trace memory model. *Psychological Review, 95,* 528-551.

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology, 33,* 36-67.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review, 89,* 609-626.

Murdock, B. B. (1983). A distributed model for serial-order information. *Psychological Review, 90,* 316-338.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). Numerical recipes in FORTRAN: The art of scientific computing (2nd ed.). New York: Cambridge University Press.

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93-134.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving from memory. *Psychonomic Bulletin & Review, 4,* 145-166.

Thompson, W. F. (1994). Sensitivity to combinations of musical parameters: Pitch with duration, and pitch pattern with durational pattern. *Perception & Psychophysics, 56,* 363-374.

Thompson, W. F., Hall, M. D., and Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 128-140.

Thompson, W. F., & Sinclair, D. (1993). Pitch pattern, durational pattern, and timbre: A study of the perceptual integration of auditory qualities. *Psychomusicology, 12,* 3-21.

Treisman, A. (1986). Properties, parts, and objects. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. II. Cognitive processes and performance* (pp. 35-70). Chichester, UK: Wiley.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136.

Treisman, A., & Paterson, R. (1984). Emergent features, attention, and object perception. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 12-31.

Treisman, A., & Schmidt, N. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107-141.

Sommaire

Les travaux de recherche actuels se servent du modèle d'erreurs pour considérer comme des échecs du processus auditif d'intégration des caractéristiques, les combinaisons de caractéristiques auditives et, par le fait même, pour expliquer ce processus. Nous proposons une autre interprétation possible de ces effets, qui repose sur les relations de similarité qui existent entre la sonde et l'ensemble de stimuli à laquelle elle est comparée.

Par exemple, dans les expériences menées par Thompson, Hall et Pressing (2001), les auditeurs ont étudié des ensembles de signaux sonores dont deux des attributs de stimulus variaient : la hauteur tonale et la durée. La présentation de l'ensemble de stimuli était suivie par un stimulus-sonde et les auditeurs devaient indiquer si la sonde se retrouvait dans l'ensemble de stimuli. Une sonde positive avait, avec l'ensemble des stimuli, trois caractéristiques communes : la hauteur tonale, la durée et la présence conjuguée de la hauteur tonale et de la durée. La hauteur tonale et la durée d'une sonde négative sélectionnable concordaient avec les caractéristiques présentées dans l'ensemble de stimuli, à l'inverse de la présence conjuguée de la hauteur tonale et de la durée. Une caractéristique de la sonde présentant un seul attribut (soit la hauteur tonale, soit la durée) concordait avec une caractéristique présentée dans l'ensemble de stimuli, tandis que pour l'autre caractéristique, cet effet n'était pas observé. On n'a observé aucune caractéristique commune entre la sonde qui présentait deux attributs et l'ensemble de stimuli.

Parce que la similarité facilite la reconnaissance d'une sonde positive et empêche le rejet d'une sonde négative, les structures de similarité des expériences permettent de prévoir le modèle de résultats prédit par la théorie de l'intégration des caractéristiques : p(correspondance positive) > p(correspondance | permutation) > p(correspondance | simple) > p(correspondance | double). Nous avons étayé notre hypothèse en appliquant le modèle REM (Shiffrin et Steyvers, 1997), un modèle général de la mémoire de reconnaissance utilisé actuellement, aux tâches de Thompson et de ses collaborateurs. Nos simulations ont démontré que la performance de reconnaissance, comme la théorie de l'intégration des caractéristiques nous permettait de le prédire, découle des relations de similarité qui existent entre la sonde et l'ensemble de stimuli à laquelle elle appartient.

Il reste néanmoins à déterminer jusqu'où nos observations peuvent s'appliquer à d'autres travaux de recherche sur l'intégration des caractéristiques dans le domaine de la recherche auditive (p. ex., Hall, Pastore, Acker et Huang, 2000). Toutefois, il semble évident, que, au moment d'élaborer le plan d'une expérience destinée à examiner le processus d'intégration des caractéristiques, les chercheurs doivent prendre en compte les relations de similarités qui relient une sonde et la série de stimuli à laquelle on la compare.